

# COVID-19 Case Surveillance Public Use Data Utility Summary

Users should consider the level of completeness, including suppression levels when planning their analyses and use of public datasets. Privacy protections will suppress field values to reduce reidentification risks. Completeness varies by jurisdiction (i.e., state, local, and territorial) and time period. Variables are consistently coded to the value “Unknown” when jurisdictions specify in the case data submitted to CDC that the value is unknown, the value “Missing” when jurisdictions do not provide a value, and the value “NA” when the value is suppressed as part of privacy protections.

Dataset version: 2021-07-06

Total records in dataset: 27,145,726

## Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	27,145,726	27,145,726
total_cells	325,748,712	81,437,178
missing_fields	66,632,188	925,868
missing_pct	20%	1%
complete_fields	259,116,524	80,511,310
complete_pct	80%	99%
unknown_fields	22,562,086	9,537,921
unknown_pct	7%	12%
suppressed_fields	112	112
suppressed_pct	0%	0%
available_fields	236,554,326	70,973,277
available_pct	73%	87%

## Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	16	0.0%	684,953	2.5%
sex	22	0.0%	55,244	0.2%
age_group	74	0.0%	185,671	0.7%
records_with_any_field	74	0.0%	865,654	3.2%